

AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently amended) A system that facilitates modeling unobserved speech dynamics, comprising:
an input component that receives acoustic data; and
a model component that employs models speech-based, at least in part, upon the acoustic data to characterize speech, the model component comprising model parameters which characterize aspects of the unobserved dynamics in speech articulation, and, which that form characterize a mapping relationship from [[the]] unobserved speech dynamics dynamic variables to observed speech acoustics, the model parameters are employed to modified-based, at least in part, upon a variational learning technique, and a technique for decoding decode an underlying unobserved phone sequence of speech based, at least in part, upon a variational learning technique,
wherein the model component is based, at least in part, upon a hidden dynamic model in the form of a segmental switching state space model.
2. (Original) The system of claim 1, modification of at least one of the model parameters being based upon a variational expectation maximization algorithm having an E-step and M-step.
3. (Original) The system of claim 2, modification of at least one of the model parameters being based, at least in part, upon a mixture of Gaussian (MOG) posteriors based on a variational technique.

4. (Currently amended) The system of claim 3, the model component being based, at least in part, upon:

$$q(s_{EN}, x_{EN}) = \prod_n q(x_n | s_n) q(s_n)$$

where x is a state of the model,

s is a phone index,

n is a frame number, [[and,]]

N is the number of frames to be analyzed, and

q is a probability approximation.

5. (Original) The system of claim 2, modification of at least one of the model parameters being based, at least in part, upon a mixture of hidden Markov model (HMM) posteriors based on a variational technique.

6. (Currently amended) The system of claim 1, the model component selecting an approximate posterior distribution relating to the acoustic data and optimizing a posterior distribution by minimizing a Kullback-Leibler [[Kullback-Liebler]] (KB) distance thereof to an exact posterior distribution.

7. (Cancelled)

8. (Currently amended) The system of claim [[7]] 1, the model component being based, at least in part, upon a switching state-space model for speech applications.

9. (Currently amended) The system of claim [[7]] 1, the model component employing, at least in part, the state equation:

$$\mathbf{x}_n = \mathbf{A}_s \mathbf{x}_{n-1} + (\mathbf{I} - \mathbf{A}_s) \mathbf{u}_s + \mathbf{w},$$

and the observation equation:

$$\mathbf{y}_n = \mathbf{C}_s \mathbf{x}_n + \mathbf{c}_s + \mathbf{v},$$

where n is a frame number,

s is a phone index,

x is the hidden dynamics,

y is an acoustic feature vector,

v is Gaussian white noise,

w is Gaussian white noise, [[and,]]

A is a phone dependent system matrix,

I is an identity matrix,

u is a target vector, and

C and c are the parameters for mapping from x to y.

10. (Currently amended) The system of claim [[7]] 1, the model component being expressed, at least in part, in terms of probability distributions:

$$p(s_n = s | s_{n-1} = s') = \pi_{ss'},$$

$$p(x_n | s_n = s, x_{n-1}) = N(x_n | \mathbf{A}_s \mathbf{x}_{n-1} + \mathbf{a}_s, \mathbf{B}_s),$$

$$p(y_n | s_n = s, x_n) = N(y_n | \mathbf{C}_s \mathbf{x}_n + \mathbf{c}_s, \mathbf{D}_s),$$

where $\pi_{ss'}$ is a phone transition probability matrix, $\mathbf{a}_s = (\mathbf{I} - \mathbf{A}_s)\mathbf{u}_s$, where A_x is a phone dependent system matrix, I is an identity matrix, and u is a target vector,

N denotes a Gaussian distribution with mean and precision matrix as the parameters,

A and a are the parameters for mapping from a state of x at a given frame to a state of x at an immediately following frame,

B represents the covariance matrix of the residual vector after the mapping from a state of x at a given frame to a state of x at an immediately following frame,

C and c are the parameters for mapping from x to y, and,

D represents the covariance matrix of the residual vector after the mapping from x to y.

11. (Original) A speech recognition system employing the system of claim 1.
12. (Currently amended) A method that facilitates modeling speech dynamics comprising:

~~recovering decoding an unobserved phone sequence of speech from acoustic data based, at least in part, upon a speech model, the speech model based upon a hidden dynamic model in the form of a segmental switching state space model and comprising having at least two sets of parameters, a first set of model parameters describing unobserved speech dynamics and a second set of model parameters describing a relationship between [[the]] an unobserved speech dynamic vector and an observed acoustic feature vector;~~

~~calculating a posterior distribution based on at least the [[above]] first set of model parameters and the second set of model parameters; and,~~

~~modifying at least one of the model parameters based, at least in part, upon the calculated posterior distribution.~~
13. (Currently amended) The method of claim 12 further comprising receiving [[the]] acoustic data.
14. (Currently amended) A method that facilitates modeling speech dynamics comprising:

~~recovering a phone sequence of speech from acoustic data based, at least in part, upon a speech model, wherein the speech model is a in the form of segmental switching state space model and comprises a plurality of model parameters;~~

~~calculating an approximation of a posterior distribution based on the model parameters, the model parameters and the approximation based upon a mixture of Gaussians; and,~~

~~modifying at least one of the model parameter[[s]] based, at least in part, upon the calculated approximated posterior distribution and minimization of a Kullback-Leibler [[Kullback-Liebler]] distance of the approximation from an exact posterior distribution.~~

15. (Currently amended) The method of claim 14 further comprising receiving [[the]] acoustic data.

16. (Currently amended) The method of claim 14, calculation of the approximation of the posterior distribution being based, at least in part, upon:

$$q(s_{EN}, x_{EN}) = \prod_n q(x_n | s_n) q(s_n),$$

where x is a state of the model,

s is a phone index,

n is a frame number, [[and,]]

N is the number of frames to be analyzed, and

q is a posterior probability approximation.

17. (Currently amended) A method that facilitates modeling speech dynamics comprising:

recovering a phone sequence of speech from acoustic data based, at least in part, upon a speech model in the form of a segmental switching state space model;

calculating an approximation of a posterior distribution based on model parameters, the model parameters and the approximation based upon a hidden Markov model posterior[[s]]; and,

modifying at least one of the model parameters based, at least in part, upon the calculated approximated posterior distribution and minimization of a Kullback-Leibler [[Kullback-Liebler]] distance of the approximation from an exact posterior distribution.

18. (Currently amended) The method of claim 17, calculation of the approximation of the posterior distribution being based, at least in part, upon:

$$q(s_{EN}, x_{EN}) = \prod_{n=1}^N q(x_n | s_n) \cdot \prod_{n=2}^N q(s_n | s_{n-1}) \cdot q(s_1)$$

where x is a state of the model,

s is a phone index,

n is a frame number, [[and,]]

N is the number of frames to be analyzed, and

q is a posterior probability approximation.

19. (Currently amended) A data packet transmitted between two or more computer components that facilitates modeling of speech dynamics, the data packet comprising:

a data structure associated with one or more recovered speech parameters; and[[,]]
the recovered speech being based, at least in part,

a segmental switching state space speech model that employs based upon acoustic data and the one or more recovered speech parameters to facilitate modeling of speech dynamics, model parameters, and the model the recovered speech parameters including at least one articulation parameter and at least one duration parameter.

20. (Currently amended) A computer readable medium ~~storage containing~~ computer executable components of a system that facilitates instructions operable to perform a method of modeling speech dynamics comprising:

receiving an input component that receives acoustic data; [[and,]]
modeling speech based on a segmental switching state space model comprising a
model component that models speech based, at least in part, upon the acoustic data, the
model component comprising model parameters including at least two sets of parameters,
a first set of parameters that describe unobserved speech dynamics and a second set of
parameters that describe a relationship between the unobserved speech dynamic vector
and an observed acoustic feature vector, and,

modifying at least one of the first set of parameters and the second set of
parameters the model parameters are modified based, at least in part, upon a variational
learning technique.

21. (Currently amended) A system that facilitates modeling speech dynamics comprising:

means for receiving acoustic data; and,
means for ~~modeling characterizing~~ speech as a segmental switching state space
model based, at least in part, upon the acoustic data,
wherein the means for modeling speech ~~employing employs~~ model parameters
that are parameters, the model parameters modified based, at least in part, upon a
variational learning technique.